

Het vraaggesprek onderzocht

Wat zegt het verloop van de interactie in survey-interviews over de kwaliteit van de vraagformulering?

Hans van der Zouwen en Wil Dijkstra

1 De behoefte aan standaardisatie in survey-interviews

De kernvariabelen van opinie- en attitudeonderzoek zijn niet rechtstreeks waarneembaar: de onderzoeker weet nooit zeker wat de ondervraagden werkelijk denken, voelen of geloven. Onderzoekers van dergelijke 'subjectieve verschijnselen' (Turner en Martin 1984) zijn daarom afhankelijk van een redeneerwijze die in de systeemmethodologie wordt aangeduid met 'black box'-analyse: men kan de binnenkant van een black box nooit waarnemen; maar door de input en de output van de black box met elkaar te vergelijken, kan men wel schattingen maken omtrent de inhoud ervan (Van der Zouwen 1974, 1982; Dijkstra en Van der Zouwen 1977).

Door mensen vragen te stellen, dus een bepaalde input te geven, en te luisteren naar de output, de antwoorden, kan een onderzoeker, althans onder bepaalde aannamen, geldige conclusies trekken omtrent de meningen of houdingen die de gestelde vragen beogen te meten.

Eén van die aannamen is dat *als* de inputs van verschillende black boxes gelijk zijn, maar hun outputs verschillen, dat dat alleen maar kan komen doordat hun inhoud verschillen. Kortom, verschillende antwoorden op dezelfde vraag duiden op verschillen in meningen; gelijke antwoorden wijzen op overeenkomst in mening. Een fundamentele voorwaarde voor opinie- en attitudeonderzoek is dan ook dat de inputs voor alle respondenten dezelfde zijn; anders kan men verschillen tussen de gegeven antwoorden niet meer eenduidig interpreteren. Vandaar de sterke behoefte aan standaardisatie in het surveyonderzoek naar 'subjectieve verschijnselen' (Beatty 1995).

Echter, methodologisch onderzoek heeft aangetoond dat volledige standaardisatie in de praktijk van het surveyonderzoek niet haalbaar is. Interviewers en ook interviewsituaties verschillen van elkaar en die verschillen blijken samen te hangen met de verkregen antwoorden. Dergelijke onbedoelde variantiebronnen, door Sudman en Bradburn (1974) 'response effecten' genoemd, kunnen de kwaliteit van de data aantasten. Bovendien is uit onderzoek van Belson (1981) gebleken dat mogelijk de *tekst* van de gestelde vraag wel voor alle respondenten gelijk is, maar niet de *betekenis* die zij daaraan hechten. De input, de vraag, lijkt dan identiek,

maar is dat feitelijk niet. En ook als interviewers strikte instructies krijgen om zich op een bepaalde manier te gedragen, dan blijkt in de praktijk toch dat zij soms gedrag vertonen dat in strijd is met die instructies: de vraagformulering wordt gewijzigd (Brenner 1982), antwoordalternatieven worden onvolledig opgesomd, of de doorvragen worden suggestief gesteld (Dijkstra en Van der Zouwen 1988; Smit 1995).

Deze uitkomsten hebben geleid tot vervolgonderzoek naar wat er werkelijk plaatsvindt tijdens het proces van dataverzameling en naar de factoren die dat proces beïnvloeden. Dergelijk 'methodenonderzoek' kan verschillende vormen aannemen. Er worden bijvoorbeeld 'split ballot'-experimenten uitgevoerd met alternatieve vraagformuleringen. Of respondenten worden, in zogenaamde cognitieve interviews (Schwarz en Sudman 1996), ondervraagd over hoe zij de gestelde vraag opgevat hebben en waarom ze voor een bepaald antwoord hebben gekozen. Of interviewers wordt, na afloop van het interview, verzocht hun indrukken omtrent het gesprek te noteren.

2 De analyse van interacties in interviews

De intensiefste vorm van methodenonderzoek op dit gebied is dat men het *vraag-gesprek* zelf gaat onderzoeken: welke interacties hebben er precies plaatsgevonden tussen de respondent en de interviewer? Zo'n interactieanalyse vereist dat het interview integraal wordt geregistreerd, bij voorkeur op videoband. Uit praktische overwegingen beperkt men zich meestal tot integrale opname op geluidsband. Om de analyse van dit materiaal overzichtelijk te houden, worden de tapes vervolgens getranscribeerd tot 'verbatim' gespreksprotocollen. Als men die protocollen ook kwantitatief wil analyseren, worden de protocollen vervolgens gecodeerd. Daarbij ontstaan bestanden waarin het geluid op de tape, het transcript van de gesproken tekst en de sequentie van de op een vraag volgende gecodeerde taalhandelingen (de vraag-antwoord sequentie), aan elkaar gekoppeld zijn (zie Dijkstra 1994 en Dijkstra te verschijnen).

Met behulp van dit materiaal kan een onderzoeker een gedetailleerde analyse uitvoeren van het verloop van de interactie, van de kenmerken van interviewer, respondent en interviewsituatie welke deze interactie mogelijk beïnvloed hebben en van het effect dat het verloop van de interactie mogelijk gehad heeft op de door de respondenten gegeven antwoorden, respectievelijk de aan hem of haar toegekende antwoordscores.

Een dergelijke analyse kan de surveyonderzoeker informeren omtrent de problemen die de deelnemers aan het interview hebben om hun taak adequaat uit te voeren (Schaeffer en Maynard 1996). Bovendien laat zo'n analyse zien hoe de deelnemers, in het bijzonder de interviewer, proberen deze problemen te overwinnen, bijvoorbeeld door de respondenten aan te sporen en door allerlei checks en

correctiewerkzaamheden uit te voeren (Sykes en Collins 1992; Houtkoop-Steenstra 1995, 1996; Van der Zouwen en Dijkstra 1995, 1996). En men kan via zo'n interactieanalyse te weten komen waar en hoe interviewers fouten maken (Dijkstra en Van der Zouwen 1988), fouten die niet zichtbaar worden bij het lezen van de ingevulde vragenlijsten, noch bij de *debriefing* van interviewers, noch bij her-interviews van respondenten.

3 De onderzoeksvragen

In dit artikel beperken wij ons tot één bepaalde toepassing van de interactieanalyse, namelijk als middel om de kwaliteit van vragen uit vragenlijsten voor survey-onderzoek te beoordelen: wat zegt het verloop van de interactie in survey-interviews over de kwaliteit van de vraagformulering?

Deze vraag berust op de veronderstelling dat er een zekere samenhang bestaat tussen de wijze waarop een vraag van een vragenlijst geformuleerd is en de interactie die op die vraag volgt. Die veronderstelling wordt in het eerste deel van dit artikel getoetst met data die afkomstig zijn uit een 'face to face'-interview onder Nederlandse jongvolwassenen naar opvattingen en plannen inzake wonen, relaties, kinderen krijgen en werkloosheid.

Daarbij zal blijken dat bepaalde kenmerken van de vraag in hoge mate van invloed zijn op de aard van de interactie. Op basis van dat onderzoeksresultaat wordt in het vervolg van dit artikel de causale relatie tussen vraag en interactie als het ware 'in omgekeerde richting' onderzocht: wat zegt het verloop van de interactie over de vraag die aan het begin van die interactie stond? En komen conclusies omtrent de kwaliteit van de vragen op basis van interactieanalyse overeen met conclusies op basis van andere methoden voor het beoordelen van vragen, zoals het analyseren van de cognitieve moeilijkheid van het beantwoorden van de gestelde vraag, of het laten beoordelen van de kwaliteit van de vragen door experts? Deze tweede onderzoeksvraag wordt onderzocht met behulp van gegevens die afkomstig zijn van een telefonisch onderzoek naar de houding van volwassen Nederlanders tegenover reclameboodschappen in de media.

4 Typen sequenties

Om de eerste onderzoeksvraag te kunnen beantwoorden zijn 1227 vraag-antwoord sequenties geanalyseerd die afkomstig waren van delen van interviews die in 1987 gehouden zijn met Nederlandse jongvolwassenen.¹ De uit deze interviews geselecteerde delen hebben betrekking op negen gesloten vragen over onderwerpen als het buitenshuis gaan wonen, meningen omtrent werkloos worden, de beoordeling van de opleiding die men gevolgd heeft en verwachtingen inzake het krijgen van kinderen.

De geluidsopnames van de geselecteerde delen van de interviews zijn getranscribeerd tot verbatim protocollen en vervolgens per vraag opgedeeld in series verbale uitingen, die elk afzonderlijk zijn gecodeerd met behulp van een computerondersteunde codeerprocedure (Sequence Viewer). Vervolgens is het computerprogramma Sequence (Dijkstra 1994) gebruikt om die 1227 vraag-antwoord sequenties onder te brengen in een typologie.

De typologie die hier gebruikt wordt, bestaat uit drie hoofdtypen:

- (1) *Paradigmatische*² sequenties, vraag-antwoord sequenties waarvan het verloop niet substantieel verschilt van de sequentie zoals die beoogd is door de onderzoeker: de vraag wordt correct voorgelezen en alle antwoordalternatieven worden opgenoemd, de respondent kiest daaruit een antwoordalternatief en de interviewer noteert dat alternatief, eventueel gevolgd door een korte reactie van de interviewer ('dank u wel').
- (2) *Problematische* sequenties, waarin weliswaar afwijkingen van de paradigmatische sequentie voorkomen, maar die adequaat 'hersteld' worden door een van beide deelnemers (meestal de interviewer), en
- (3) *Inadequate* sequenties waarin dit herstel niet plaatsvindt, althans niet op adequate wijze.

5 Vragen en het optreden van problemen en onvolkomenheden

De verdeling van de paradigmatische, problematische en inadequate sequenties verschilt sterk voor de negen gekozen vragen. Het percentage problematische sequenties varieert van 2 tot 34% en het percentage inadequate sequenties varieert van 3 tot 46%. In tabel 1 zijn de negen vragen gerangschikt naar opklimmend percentage inadequate sequenties.

Drie vragen (V146, V46 en V41) veroorzaken weinig problemen en leiden slechts in enkele gevallen tot inadequate sequenties. Deze drie vragen nodigen de respondenten uit om hun (on)tevredenheid met een bepaald aspect van het leven (opleiding, werkkring of relatie) tot uitdrukking te brengen door het kiezen van een van de alternatieven van toonkaart 11. De meeste sequenties die op deze vragen volgen, vertonen hetzelfde, paradigmatische, patroon: de vraag wordt correct voorgelezen door de interviewer en de respondent noemt een van de alternatieven van de aangeboden toonkaart, soms gevolgd door een korte toelichting op deze keuze en een reactie van de interviewer.

De vragen V97, V152 en V99 leiden tot een hoger percentage van zowel problematische als inadequate sequenties. Voor de drie resterende vragen is het percentage inadequate sequenties zelfs verontrustend hoog.

Tabel 1: Percentage paradigmatische, problematische en inadequate vraag-antwoord sequenties volgend op negen vragen uit het 'face to face'-survey

| Vraag | Paradig matisch | Proble matisch | Inade quaat | N |
|--|--------------------|-------------------|----------------|-----|
| V146 Hoe vind je het dat je geen vaste relatie hebt? Hoe tevreden of ontevreden ben je daarmee? Int. Gebruik kaart 11 (zie V41) | 95% | 2% | 3% | 66 |
| V46 Hoe tevreden ben je al met al met de baan die je momenteel hebt? Int. Gebruik kaart 11 (zie V41) | 92% | 5% | 3% | 100 |
| V41 Hoe tevreden ben (was) je al met al over de opleiding die je volgt (gevolgd hebt)? Int. Gebruik kaart 11 : 1. heel ontevreden; 2. tamelijk ontevreden; 3. noch tevreden, noch ontevreden; 4. tamelijk tevreden; 5. heel tevreden. | 85% | 8% | 7% | 181 |
| V99 Zal je met het leven alles bij elkaar genomen meer of minder tevreden worden als je werkloos wordt (blijft)? Int. Gebruik kaart 9 : 1. veel minder tevreden; 2. iets minder tevreden; 3. ongeveer gelijk; 4. iets meer tevreden; 5. veel meer tevreden. | 76% | 14% | 10% | 171 |
| V97 Hoe lijkt het je om werkloos te worden (blijven)? 1. zeer onaangenaam; 2. onaangenaam; 3. niet aangenaam en niet onaangenaam; 4. aangenaam; 5. zeer aangenaam. | 60% | 24% | 16% | 171 |
| V152 Ben je van plan in de toekomst (nog een of meer) kinderen te krijgen? 1. beslist niet; 2. waarschijnlijk niet; 3. weet (nog) niet; 4. waarschijnlijk wel; 5. zeker wel. | 46% | 34% | 20% | 172 |
| V96 Heb je er wel eens over nagedacht hoe het zou zijn als je onvrijwillig werkloos zou worden (blijven)? 1. haast nooit; 2. weinig; 3. af en toe; 4. vaak; 5. heel vaak. | 53% | 18% | 29% | 170 |
| V22 Hoe moeilijk denk je dat het voor jou is om betaalbare woonruimte te vinden, als je het huis uit zou gaan? 1. zal geen enkel probleem opleveren; 2. zal weinig moeite kosten; 3. zal tamelijk veel moeite kosten; 4. zal bijzonder veel moeite kosten; 5. zal bijna onmogelijk zijn. | 41% | 19% | 40% | 98 |
| V21 Hoeveel moeite doe je om woonruimte te vinden? 1. geen enkele moeite; 2. een beetje moeite; 3. veel moeite | 35% | 19% | 46% | 98 |

6 Vraagmoeilijkheid en inadequaat verlopende sequenties

Uit tabel 1 blijkt dus dat er een sterke samenhang bestaat tussen de vraag en de mate van adequaatheid van de op die vraag volgende interactie. Maar welke kenmerken van de vraag zijn nu verantwoordelijk voor het al dan niet adequaat verlopen van de interactie? Om die vraag te kunnen beantwoorden, is een globale ana-

lyse gemaakt van de inhoud van de vraag en de daarbij behorende antwoordalternatieven. Het achterliggende idee is dat bepaalde kenmerken van een vraag haar moeilijker te beantwoorden maken, hetgeen kan leiden tot een toename van inadequaate verlopende sequenties.

Ten behoeve van deze analyse onderscheidden we in eerste instantie vier componenten van 'vraagmoeilijkheid':

1. *Hypothetische of voorwaardelijke vraag*: Een voorbeeld van een voorwaardelijke vraag is vraag V22: 'Hoe moeilijk denk je dat het voor jou is om betaalbare woonruimte te vinden, *als je het huis uit zou gaan?*'

2. *Ambigüiteit van de vraag*: Vraag V96 luidt: 'Heb je er wel eens over nagedacht hoe het zou zijn als je onvrijwillig werkloos zou worden?' Uit de antwoorden die de respondenten op deze vraag geven, blijkt dat sommigen de nadruk leggen op het eerste deel van de vraag (hoe vaak men erover *gedacht* heeft), terwijl anderen reageren op het tweede deel van de vraag (welke *gevoelens* werkloosheid bij hen oproept). Anders gezegd, respondenten reageren bij vraag V96 eigenlijk op twee verschillende vragen.

3. *Inadequaetheit van de aangeboden antwoordalternatieven*: Ook op dit punt scoort vraag V96 slecht: De 'officiële' antwoordalternatieven zijn: 1. haast nooit; 2. weinig; 3. af en toe; 4. vaak; 5. heel vaak. Dat is vreemd, want formeel gezien is vraag V96 een ja-nee vraag, terwijl de antwoordalternatieven een bepaalde frequentie uitdrukken. Bovendien ontbreekt het alternatief 'nooit', terwijl dit juist een alternatief is dat door nogal wat respondenten spontaan genoemd wordt. En ten slotte, het onderscheid tussen de antwoordalternatieven 'weinig' en 'af en toe' is wel zeer gering. De gebrekkige afstemming tussen vraagtekst en antwoordalternatieven leidt dan ook tot heel wat 'extra werk' voor interviewer en respondent.

4. *Afwezigheid van een toonkaart*: Door Van der Zouwen en Dijkstra (1996) is er al op gewezen dat het gebruik van toonkaarten met daarop de antwoordalternatieven waaruit de respondent een keuze kan maken, een bijzonder heilzame invloed heeft op het verloop van de interactie. De vier vragen waarbij een toonkaart gebruikt is, zijn de vragen met het laagste percentage inadequate sequenties.

In tabel 2 staat voor elk van de negen vragen aangegeven in hoeverre een component van taakmoeilijkheid niet (0) of wel (1) aanwezig is en wat het percentage inadequaate verlopende sequenties voor elk van deze negen vragen is. Uit de rij met correlatiecoëfficiënten blijkt dat elk van de vier componenten positief samenhangt met het percentage inadequate sequenties. Een multipele regressieanalyse toont dat de derde (alternatieven) en vierde (toonkaart) component van taakmoeilijkheid 87% van de variantie in de afhankelijke variabele verklaren (multiple $R = .93$), een uitkomst die zeer significant is. Deze bevindingen geven een antwoord op onze eerste onderzoeksvraag: het zijn vooral de adequaatheid van de alternatieven en de wijze waarop ze aan de respondent worden aangeboden die het verloop van de interactie bepalen.

Tabel 2: Componenten van taakmoeilijkheid en percentage inadequate sequenties

| Vraag | Condi- tionele vraag | Ambi- gue vraag | Inade- quate reeks alterna- tieven | Geen toon- kaart ge- bruikt | % Inade- quate sequen- ties |
|--|----------------------------|-----------------------|--|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| V146 Hoe vind je het dat je geen vaste relatie hebt? Hoe tevreden of ontevreden ben je daarmee? | 0 | 0 | 0 | 0 | 3% |
| V46 Hoe tevreden ben je al met al met de baan die je momenteel hebt? | 0 | 0 | 0 | 0 | 3% |
| V41 Hoe tevreden ben (was) je al met al over de opleiding die je volgt (gevolgd hebt)? | 0 | 0 | 0 | 0 | 7% |
| V99 Zal je met het leven alles bij elkaar genomen meer of minder tevreden worden als je werkloos wordt (blijft)? | 1 | 0 | 1 | 0 | 10% |
| V97 Hoe lijkt het je om werkloos te worden (blijven)? | 1 | 0 | 0 | 1 | 16% |
| V152 Ben je van plan in de toekomst (nog een of meer) kinderen te krijgen? | 1 | 0 | 0 | 1 | 20% |
| V96 Heb je er wel eens over nagedacht hoe het zou zijn als je onvrijwillig werkloos zou worden (blijven)? | 1 | 1 | 1 | 1 | 29% |
| V22 Hoe moeilijk denk je dat het voor jou is om betaalbare woonruimte te vinden, als je het huis uit zou gaan? | 1 | 0 | 1 | 1 | 40% |
| V21 Hoeveel moeite doe je om woonruimte te vinden? | 0 | 1 | 1 | 1 | 46% |
| Correlatiecoëfficiënt | .27 | .65 | .71 | .81 | |

Eerder was door Smit (1995) al geconstateerd dat een bij surveyonderzoekers betrekkelijk populair type vragen bijzonder veel inadequate sequenties oplevert. Het betreft zogenaamde semi-open vragen ofwel *field-coded questions* (Sudman en Bradburn 1982). Smit (1995, p. 79) definieert semi-open vragen als

‘vragen met een open antwoordformaat, waarbij een lijst met antwoordalternatieven is gevoegd. In principe mogen de interviewers deze alternatieven echter niet aan de respondenten voorleggen. Het is de bedoeling dat de interviewers met behulp van de open vraag antwoorden aan de respondenten ontlokken die vervolgens door middel van “field coding” worden gescoord met behulp van deze lijst antwoordalternatieven’.

Aan de jongvolwassenen, die deelnamen aan het ‘face to face’-survey waaruit de hiervoor besproken data afkomstig zijn, werd bijvoorbeeld de open vraag gesteld (V20) ‘Om wat voor soort reden zou je eventueel uit huis willen gaan?’ Op het formulier van de vragenlijst zelf waren tien verschillende redenen afgedrukt, maar

die mochten door de interviewer niet opgelezen worden. Bij de interactieanalyse blijkt dat dit vraagformaat tot veel inadequaat interviewgedrag leidt: het – tegen de instructie in – toch oplezen van de alternatieven, het onvoldoende doorvragen als het gegeven antwoord niet past bij een van de alternatieven, of het, in reactie op een niet scorebaar antwoord van de respondent, suggestief aanbieden van een kandidaat-alternatief. Smit (1995, p. 88) concludeert dan ook: 'De kwaliteit van data verkregen met semi-open vragen is niet goed'.

Op basis van de tot nu toe gepresenteerde onderzoeksresultaten komen we tot vier praktische aanbevelingen:

- (1) indien men vragen uit (concept)vragenlijsten gaat beoordelen, dient men bij die beoordeling ook de aangeboden antwoordalternatieven te betrekken;
- (2) vermijd het gebruik van semi-open vragen;
- (3) gebruik (bij een groter aantal alternatieven) zoveel mogelijk toonkaarten;
- (4) in die situaties waarin het gebruik van toonkaarten niet of nauwelijks mogelijk is, zoals in telefonische interviews, dient de formulering van de vraag en de keuze van de antwoordalternatieven met (nog meer) zorg te geschieden.

7 Het opsporen van problematische sequenties in het CATI-survey

Onze tweede onderzoeksvraag luidt: tot welke aanwijzingen leidt een analyse van de interactie omtrent het problematisch, of zelfs inadequaat zijn van vragen; en zijn deze aanwijzingen gelijk of verschillend aan die welke gebaseerd zijn op een beoordeling van vragen door een jury van experts, of op een analyse van de cognitieve moeilijkheid van het beantwoorden van die vragen?

Om die vraag te beantwoorden maken we gebruik van een tweede dataset, die bestaat uit 1943 sequenties afkomstig van 54 computergestuurde telefonische interviews die eind 1995 en begin 1996 gehouden zijn met Nederlandse volwassenen.³ De 37 uit deze interviews geselecteerde vragen gaan met name over tv-kijkgedrag en over meningen en houdingen ten opzichte van tv-reclame.

Voor elk van die 37 vragen hebben we (1) een beoordeling gemaakt van de cognitieve moeilijkheid van het beantwoorden van die vraag, (2) een jury van experts op het gebied van surveyonderzoek een beoordeling laten geven van de kwaliteit van de formulering van de vraag, en (3) de vraag-antwoord sequenties die tot deze vragen behoren, geanalyseerd en wel door het tellen van het aantal afwijkingen van de paradigmatische sequentie.

Met betrekking tot de moeilijkheid van het beantwoorden van een surveyvraag worden twee componenten onderscheiden: de moeilijkheid van de vraag zelf, en de onduidelijkheid van de taak.⁴

De moeilijkheid van een vraag zal (ceteris paribus) groot zijn als:

- (1) er veel cognitieve operaties uitgevoerd moeten worden om tot een antwoord te komen;
- (2) het noodzakelijk is om informatie uit het lange termijn geheugen op te halen;
- (3) het voor het adequaat beantwoorden van de vraag noodzakelijk is om informatie omtrent de vraag en de antwoordalternatieven op te slaan in het korte termijn geheugen;
- (4) de vraag verwijst naar een hypothetische of toekomstige situatie;
- (5) (voor open vragen) er veel precisie van het antwoord geëist wordt;
- (6) (voor gesloten vragen) de antwoordalternatieven slecht passen bij de vraag zelf;
- (7) (voor gesloten vragen) de antwoordalternatieven ongelijk verdeeld zijn over de *range* van mogelijke waarden van de variabele die met deze vraag gemeten wordt;
- (8) geen gebruikgemaakt wordt van hulpmiddelen om het antwoorden te vereenvoudigen zoals toonkaarten, plaatjes, ladders en barometers;
- (9) het onderwerp van de vraag leidt tot problemen op het gebied van 'zelfpresentatie' (bedreigendheid van het onderwerp; grote verschillen inzake de sociale wenselijkheid van de antwoordalternatieven).

De onduidelijkheid van de taak zal groot zijn als:

- (10) de formulering van de vraag ambigu is;
- (11) de vraag onbekende woorden bevat;
- (12) duidelijke aanwijzingen omtrent beurtwisseling ontbreken.

De 37 vragen van de CATI-vragenlijst zijn door ons beoordeeld op elk van deze 12 indicatoren van taakmoeilijkheid. Met betrekking tot elke indicator werd nagegaan of de betrokken vraag beduidend moeilijker is dan een vraag die als referentiepunt diende en die gemakkelijk beantwoord kan worden door iedere respondent zonder verdere uitleg of aanmoediging van de zijde van de interviewer (bijvoorbeeld: 'Hoe oud bent u?'). Dat betekent dat de score voor taakmoeilijkheid in principe kan lopen van 0 tot 12 punten.

Het blijkt dat deze vragenlijst geen zeer moeilijke vragen bevat: geen van de vragen heeft een moeilijkheidsscore (TaakAnalyse- of TA-score) van 7 of meer punten.⁵ Veertien vragen hebben een TA-score van 0 tot en met 2 punten, hetgeen betekent dat ze betrekkelijk gemakkelijk te beantwoorden zijn. Achttien vragen hebben TA-scores van 3, 4 of 5 punten; deze vragen zijn niet zo gemakkelijk te beantwoorden. Slechts vijf vragen hebben een TA-score van 6 punten, hetgeen hen bestempelt als 'tamelijk moeilijk' (zie tabel 3). Vier van deze vijf vragen (V9 tot en met V12) vereisen voor de beantwoording een beroep op het lange termijn geheugen, omdat ze een situatie uit het verleden ('de laatste keer dat u tv keek') betreffen en verzoeken om een nogal precieze weergave van hun beoordeling van die

Tabel 3: Vragen uit het CATI-survey, aangewezen als 'problematisch' door de cognitieve taak analyse (TA>5; range 0-12), de jury beoordeling (JB>5.0, range 1-7) en/of de interactie analyse (IA>.60; range 0-1)

| Vraag | TA (1) | JB (2) | IA (3) |
|--|-----------|------------|------------|
| V2 En op de dagen dat u kijkt, hoeveel uur of minuten kijkt u dan gemiddeld? | 3 | 5.1 | .33 |
| V9 Hoeveel procent van de tijd heeft u met aandacht gekeken? | 6 | 5.6 | .56 |
| V10 U vertelde hiervoor hoelang u naar de televisie keek. Hoeveel reclameblokken zag u ongeveer gedurende deze periode? | 6 | 5.9 | .48 |
| V11 Hoeveel reclamespots heeft u toen ongeveer gezien? | 6 | 5.7 | .40 |
| V12 Hoeveel procent van deze reclamespotjes heeft uw aandacht getrokken? | 6 | 5.8 | .42 |
| V14 Ik neem een aantal mogelijke reacties op reclame. Wilt u aangeven of u dat elke keer, vaak, soms of nooit doet als er reclame in beeld komt. U schakelt over naar een andere zender? | 4 | 3.9 | .63 |
| V15 (idem) U zet het geluid uit? 43.5.65 V17 (idem) U doet even iets anders, maar laat de reclame aan? 63.9.43 V18 (idem) U loopt de kamer uit? | 3 | 3.5 | .63 |
| V21 Ik vind reclamespotjes op tv leuk om naar te kijken? 1. zeer mee oneens; 2. mee oneens; 3. niet mee eens/niet mee oneens; 4. mee eens; 5. zeer mee eens. | 1 | 3.2 | .69 |
| V22 Ik vind reclamespotjes op tv zinvolle informatie geven over het productgebruik van andere mensen (zie V21) | 5 | 5.1 | .59 |
| V26 Ik vind reclamespotjes op tv te schreeuwerig? (zie V21) | 1 | 4.5 | .61 |
| V32 Staat u in het algemeen negatief of positief tegenover radioreclame? 1. zeer negatief; 2. negatief; 3. neutraal; 4. positief; 5. zeer positief | 2 | 4.1 | .72 |
| V35 Staat u in het algemeen negatief of positief tegenover reclame (zie V32) | 2 | 4.4 | .65 |
| Alle 1943 sequenties tezamen | 3.2 | 4.2 | .49 |

voorbij situatie (V12: 'Hoeveel procent van deze reclamespotjes heeft uw aandacht getrokken?').

8 Jury-beoordelingen van het problematisch karakter van surveyvragen

Vragen die tot problemen leiden bij de beantwoording, kunnen de kwaliteit van de gegevens ongunstig beïnvloeden. Daarom zou het prettig zijn als een jury, bestaande uit experts op het gebied van surveys, voorafgaande aan het onderzoek dergelijke problematische vragen zou kunnen aanwijzen. Of algemener gesteld, het zou mooi zijn als een jury in staat was de 'kwaliteit' van een vraag uit de vragenlijst vooraf te beoordelen. Om de toepasbaarheid van deze gedachte te testen,

hebben wij aan een jury van tien ervaren surveyonderzoekers gevraagd om de kwaliteit van de 37 vragen uit de CATI-vragenlijst die aan alle respondenten waren voorgelegd, te beoordelen. Hen werd gevraagd iedere vraag te plaatsen op een schaal die liep van 1 ('uitstekende vraag') tot 7 ('waardeloze vraag').

Het bleek dat de mate van consensus tussen de juryleden vrij gering was: de standaardafwijking van de scores bedroeg voor sommige vragen meer dan 2 schaalpunten. Zes vragen kregen een 'probleemscore' van 4.0 of meer. Vier daarvan, de vragen V9 tot en met V12, werden ook al bij de analyse van de taakmoeilijkheid aangewezen als problematisch (met een TA-score van 6 punten). De beide andere vragen, V2 en V22, werden bij de cognitieve taakanalyse aangemerkt als 'niet gemakkelijk te beantwoorden' vragen.

Om na te gaan in hoeverre de scores van de cognitieve taakanalyse en die van de jurybeoordeling samenhangen, is de rangcorrelatie berekend over alle 37 vragen. Spearman's rho bedraagt .35, hetgeen aangeeft dat er een weliswaar positieve, maar niet erg sterke, samenhang bestaat tussen beide beoordelingen.

9 Afwijkingen van de paradigmatische sequentie

De beantwoording van de tweede onderzoeksvraag vereist dat we de uitkomsten van de beide *a priori* beoordelingen van de vragen vergelijken met een beoordeling van de interactie die volgt op deze vragen. Het feit dat de 1943 sequenties gecodeerd zijn, maakt het mogelijk om op vrij eenvoudige wijze in die sequenties te zoeken naar de aanwezigheid van bepaalde afwijkingen van de paradigmatische sequentie. Ten behoeve van de onderhavige interactieanalyse hebben we ons geconcentreerd op vier van dergelijke afwijkingen:

- (a) het niet correct voorlezen van de vraag door de interviewer;
- (b) het kennelijk niet begrijpen van de vraag door de respondent;
- (c) het vóórkomen van suggestief doorvragen door de interviewer; en
- (d) het onvoldoende doorvragen door de interviewer.

Sequenties waarin een of meer van dergelijke afwijkingen voorkomen zijn gecodeerd als 'problematisch'.

Bij analyse blijkt dat, hoewel de vragen uit de vragenlijst bij de beide *a priori* beoordelingen in het algemeen niet als 'moeilijk' beoordeeld zijn, ze toch gevolgd worden door heel wat afwijkingen van de paradigmatische sequentie. Voor alle 37 vragen tezamen is het aantal afwijkingen per 100 sequenties zelfs 49.

De zojuist onderscheiden vier afwijkingen zijn bijzonder scheef verdeeld over de 37 vragen (chi-kwadraat is zeer significant; $p < 0.001$). Bepaalde vragen leiden duidelijk tot een bepaald soort afwijkingen. Zeven vragen hebben meer dan 6 afwijkingen per 10 sequenties, of anders gezegd, ze hebben een interactiescore (IA) van .60.

10 Een vergelijking van de uitkomsten van de drie beoordelingen

Tabel 3 biedt een overzicht van de uitkomsten van elk van de drie gehanteerde beoordelingsprocedures. De tabel bevat alleen die vragen die als 'problematisch' zijn beoordeeld door minstens één van die drie procedures; dus een TaakAnalyse (TA)-score heeft van 6, een JuryBeoordeling (JB)-score van 5 of meer, en/of een InteractieAnalyse (IA)-score van .60 of meer. De scores die deze drempels overschrijden zijn vet gedrukt in de tabel.

Tabel 3 leidt tot de volgende conclusies:

1. Het feit dat slechts 14 van de 37 vragen in deze tabel voorkomen, betekent dat 23 vragen als betrekkelijk onproblematisch beoordeeld worden door elk van de drie gebruikte procedures.

2. Als men de eerste kolom van de tabel vergelijkt met de tweede, blijkt dat de juryleden vrijwel dezelfde vragen als problematisch aanwijzen als de vragen die bij de taakanalyse als problematisch genoemd worden.

3. Echter, er zijn grote verschillen tussen de uitkomsten van de beide *a priori* beoordelingen (TA en JB) enerzijds, en de *a posteriori* beoordeling van de interacties (IA) anderzijds. Geen van de zeven vragen die op grond van de taakanalyse en/of in de jurybeoordeling als problematisch werden aangemerkt, blijken bij de interactieanalyse op te vallen door een bovengemiddeld aantal afwijkingen van de paradigmatische sequentie. Aan de andere kant worden de zeven vragen die tot een ver boven gemiddeld aantal afwijkingen in de interacties leidden, door geen van beide *a priori* beoordelingen als 'problematisch' geïdentificeerd.

Het voorlopig antwoord op onze tweede onderzoeksvraag is dan ook dat de beide *a priori* beoordelingen ongeveer dezelfde vragen als problematisch aanmerken, *maar dat dat niet de vragen zijn die tijdens de interactie de meeste problemen blijken op te leveren*. Deze bevinding nodigt uit tot een nadere analyse van de sequenties volgend op de zeven vragen met een hoge IA-score.

11 Analyse van de zeven onverwacht problematische vragen

In tabel 4 staat voor ieder van deze zeven vragen aangegeven in welke mate elk van de vier onderscheiden afwijkingen van de paradigmatische sequentie voorkomt. Uit de interactieanalyse blijkt dat afwijking (a), het niet correct voorlezen van de vraagtekst door de interviewer, niet veel voorkomt; feitelijk alleen bij de vragen V44 en V46, maar niet bij de zeven vragen uit tabel 4. En afwijking (b), het kennelijk niet begrijpen van de vraag door de respondent, deed zich vooral voor bij, de niet in tabel 4 vermelde, vraag V6. Deze vraag ('Waar heeft u vanavond gekeken?') werd door een aantal respondenten kennelijk begrepen als 'Waarnaar

Tabel 4: Voorkomen van afwijkingen van de paradigmatische sequentie [(a) onjuist voorlezen van de vraag; (b) niet begrijpen van de vraag; (c) suggestief doorvragen; (d) onvoldoende doorvragen] per sequentie voor de zeven vragen die op grond van de interactie analyse als 'problematisch' zijn aangemerkt (IA score >.60 range 0-1)

| Vraag | (a) | (b) | (c) | (d) | Totaal |
|---|-----|-----|------------|------------|--------|
| V14 Ik noem een aantal mogelijke reacties op reclame. Wilt u aangeven of u dat elke keer, vaak, soms of nooit doet als er reclame in beeld komt? U schakelt naar een andere zender? | .02 | .04 | .22 | .35 | .63 |
| V15 (idem) U zet het geluid uit? | .00 | .02 | .17 | .46 | .65 |
| V18 (idem) U loopt de kamer uit? | .00 | .02 | .20 | .41 | .63 |
| V21 Ik vind reclamespotjes op tv leuk om naar te kijken? 1. zeer mee oneens; 2. mee oneens; 3. niet mee eens/niet mee oneens; 4. mee eens; 5. zeer mee eens. | .00 | .00 | .33 | .35 | .69 |
| V26 Ik vind reclamespotjes op tv te schreeuwerig? (zie V21) | .00 | .06 | .15 | .41 | .61 |
| V32 Staat u in het algemeen negatief of positief tegenover radioreclame? 1. zeer negatief; 2. negatief; 3. neutraal; 4. positief; 5. zeer positief. | .04 | .06 | .22 | .41 | .72 |
| V35 Staat u in het algemeen negatief of positief tegenover reclame? (zie V32) | .00 | .05 | .17 | .43 | .65 |
| Alle 1943 sequenties tezamen | .03 | .06 | .14 | .27 | .49 |

heeft u vanavond gekeken?' en overeenkomstig beantwoord door het opsommen van tv-programma's.

De afwijkingen die vooral leiden tot hoge IA-scores zijn (c) suggestief doorvragen en (d) onvoldoende doorvragen. Met name uitspraak V21 ('Ik vind reclamespotjes op tv leuk om naar te kijken') leidt tot veel suggesties van de zijde van de interviewer. Om een voorbeeld te geven: de respondent reageert op deze uitspraak met 'nee', een antwoord dat niet voorkomt in het rijtje met antwoordalternatieven en dus ook niet scorebaar is. De interviewer legt vervolgens de respondent als 'kandidaat-antwoord' het alternatief 'zeer mee oneens' voor, daarmee een suggestie aanreikend inzake de intensiteit van de mening van de respondent. Deze afwijking van de paradigmatische sequentie wordt niet veroorzaakt door een inadequate vraagformulering, maar door inadequaat gedrag van de interviewer. Als men de suggesties van de interviewers nader beziet, blijken ze allemaal in één richting te gaan: reclamespotjes zijn helemaal niet leuk en dat vinden de respondenten (dus) ook. De aldus verkregen scores op deze vraag zeggen meer over de verwachtingen van de interviewers, dan over de intensiteit van de meningen van de respondenten.

Het onvoldoende doorvragen (afwijking (d)) gebeurt vooral bij de vragen V15 en V18. Dat is opmerkelijk, want deze vragen zien er op het eerste gezicht ade-

quaat uit. Ze werden dan ook niet opgespoord bij de cognitieve taakanalyse of bij de jurybeoordeling. Pas nadat verschillende interviews waren afgeluisterd, bleek wat het probleem met deze vragen is: ze maken deel uit van een serie van zeven vragen naar de frequentie waarmee de respondent op een bepaalde wijze reageert als op het tv-scherm reclame verschijnt. Geen van deze zeven vragen is op zichzelf problematisch. Maar tezamen kunnen ze interviewer en respondent in moeilijkheden brengen. Dat komt omdat het antwoord op de ene vraag, het stellen en beantwoorden van de volgende vraag nogal lachwekkend kan maken: als de respondent bij vraag V14 gezegd heeft dat hij altijd op tv-reclame reageert door naar een ander kanaal te gaan, klinkt het noemen – bij vraag 15 – van de optie ‘het geluid uitzetten’ nogal vreemd. Dat leidt ertoe dat de interviewer geneigd is om welke reactie de respondent ook geeft op vraag V15, deze reactie maar zelf te ‘transformeren’ naar een van de vier antwoordalternatieven, ook als de door de respondent geboden informatie beslist onvoldoende is om dat te doen. In dit geval is het niet de afzonderlijke vraag die problemen oplevert, maar de combinatie van – zelfs logisch – samenhangende vragen.

De vragen V32 en V35 leiden tot onvoldoende doorvragen, omdat in telefonische interviews de interviewer geen toonkaart kan gebruiken. Dus iedere keer als de respondent de vraag ‘Staat u in het algemeen negatief of positief tegenover ...?’ met ‘negatief’ heeft beantwoord, dient de interviewer door te vragen: ‘Is dat ‘zeer negatief’ of ‘(gewoon) negatief?’’, opdat score 1 danwel score 2 kan worden toegekend. Deze procedure leidt tot ongemakkelijke interacties, dus interviewers trachten dat te voorkomen door het eerste antwoord van de respondent te herinterpreteren tot een van de vijf antwoordalternatieven, zonder de respondent daar verder over te raadplegen. Het verhaal bij vraag V26 is hetzelfde. Opnieuw kan gezegd worden dat de formulering van de vraag en van de antwoordalternatieven op zich niet inadequaat is, maar dat het ontbreken van een toonkaart leidt tot veel inefficiënte interacties of riskante interpretaties van de antwoorden, door de interviewer.

12 Conclusies

Uit de analyse van de interacties die volgen op het stellen van een vraag uit de vragenlijst van een survey-interview blijkt, dat bepaalde kenmerken van de vraag samenhangen met de kans dat die interactie ‘paradigmatisch’ verloopt, dat wil zeggen, verloopt zoals beoogd door de onderzoeker. Afwijkingen van de paradigmatische vraag-antwoord sequentie kunnen leiden tot verminderde kwaliteit van de gegevens en in ieder geval tot verminderde vergelijkbaarheid van de verkregen scores.

Indien de onderzoeker ervoor zorgt dat antwoordalternatieven goed passen bij de vraagtekst en gelijkmatig gespreid zijn over de variabele die de onderzoeker met de vraag wenst te meten, en indien de alternatieven worden aangeboden met

gebruikmaking van een toonkaart, dan kunnen al veel inadequaate verlopende interacties vermeden worden. In die zin geven de uitkomsten van dit onderzoek een empirische bevestiging van de regels inzake vragenlijstconstructie zoals onder meer geformuleerd door Foddy (1993).

Omdat in telefonische surveys de toonkaart nauwelijks gebruikt kan worden, is een zorgvuldige formulering van de vragen daar nog belangrijker. Gebleken is dat noch een beoordeling van de cognitieve moeilijkheid van de antwoordtaak, noch een beoordeling van de kwaliteit van de voorgestelde vraagformulering in staat is 'vooraf' alle vragen die leiden tot veel inadequaat interacties, op te sporen. In dit onderzoek bleek dat een serie op zich correct geformuleerde vragen problemen gaf, omdat het antwoord op de ene vraag het stellen van de andere vraag vrij lachwekkend kon maken. En dat het in twee stappen vaststellen van eerst de richting van een opvatting en vervolgens de intensiteit van die opvatting, ook tot onhandige interacties of inadequaate interviewergedrag kon leiden, omdat het eerste antwoord van de respondent vaak al een indicatie geeft van zowel de richting als de intensiteit.

Uit het vervolg van onze analyse blijkt dat de aan een vraag te onderscheiden kenmerken verschillende gevolgen hebben voor de op die vraag volgende interactie:

1. De *cognitieve moeilijkheid* van de vraag kan leiden tot langere reactietijden (*response latency*), tot het herhalen van de vraag door de respondent en tot verzoeken om verduidelijking aan de interviewer. Echter, zolang de respondent in antwoord op de vraag slechts een van de aangeboden antwoordalternatieven noemt, zal de sequentie 'paradigmatisch' zijn en dus noch de moeilijkheden die de respondent heeft met het beantwoorden, noch de fouten die gemaakt worden door het verkeerd begrijpen van de vraag of het per ongeluk kiezen van het verkeerde antwoordalternatief, 'aan het licht brengen'. Een paradigmatisch verlopende interactie is nog geen garantie voor een goede vraag.

2. *Inadequate antwoordalternatieven* leiden tot 'onderhandelingen' tussen respondent en interviewer over het alternatief dat het beste past, of tot suggesties in die richting door de interviewer, of tot onvoldoende doorvragen. Niet paradigmatisch verlopende interacties kunnen duidelijke aanwijzingen bieden voor dit type gebreken in de vraagformulering.

3. Inadequaate verlopende interacties kunnen ook het gevolg zijn van niet voorziene *context-effecten*: antwoorden op eerdere vragen kunnen bij de interviewer tot verwachtingen leiden met betrekking tot het antwoord op de voorliggende vraag. En die verwachtingen van de interviewer kunnen de oorzaak zijn van suggestieve vraagstelling en onvoldoende doorvragen en daardoor afbreuk doen aan de kwaliteit en zeker de vergelijkbaarheid van de verzamelde gegevens.

Het antwoord op de vraag 'Wat zegt het verloop van de interactie in survey-inter-

views over de kwaliteit van de vraagformulering?' kan als volgt worden samengevat: Een interactieanalyse van interviews die in een vooronderzoek zijn gehouden, met een voorlopige versie van de vragenlijst afgenomen bij een kleine deelsteekproef, kan problemen aan het licht brengen die niet herkend worden door de gebruikelijke *a priori* beoordelingen van de vragenlijst (Foddy, 1996).

Echter, interactieanalyse biedt geen volledige vervanging voor andere beoordelingsprocedures, omdat ook in paradigmatische sequenties informatie overgebracht kan worden die gebaseerd is op het verkeerd begrijpen van de vraag of die behept is met fouten die hebben plaatsgevonden bij de keuze uit de lijst met antwoordalternatieven.

Noten

Dit artikel is een bewerking van een paper dat onder dezelfde titel is gepresenteerd tijdens de Methodologendag die op 14 november 1997 in Amsterdam werd gehouden.

1. De data behorende tot dit 'face to face'-survey zijn verzameld in het kader van het onderzoeksprogramma: 'Het proces van sociale integratie van jongvolwassenen. Een psychometrisch en sociometrisch onderzoek' (SI-project). Dit onderzoeksprogramma is uitgevoerd aan de Vrije Universiteit te Amsterdam door de vakgroepen Methoden en Technieken van Sociaal Wetenschappelijk Onderzoek (faculteit SCW) en Arbeids- en Organisationspsychologie (faculteit FPP).

2. De nogal zwaarwichtige term 'paradigmatische' sequentie is afkomstig van Schaeffer en Maynard (1996) en wordt in toenemende mate gebruikt bij dit type interactieanalyse.

3. Met erkentelijkheid zij vermeld dat voor deze analyse gebruikgemaakt kon worden van gegevens die afkomstig zijn van een onderzoek dat gehouden is door Edith Smit van de Universiteit van Amsterdam.

4. Deze lijst met aspecten van vraagmoeilijkheid is gebaseerd op onderzoek naar responsefouten in surveys (bijv. Schwarz en Sudman 1996; Van der Vaart 1996).

5. Opmerkelijk is dat alle 37 vragen bepaalde vraagkenmerken gemeen hebben, ofwel dezelfde waarde hebben op een van de twaalf onderscheiden indicatoren van vraagmoeilijkheid. Zo verwijst geen van de vragen naar een hypothetische of toekomstige situatie (indicator 4) en wordt in geen van de vragen gebruikgemaakt van hulpmiddelen om het antwoorden te vereenvoudigen zoals toonkaarten, plaatjes, ladders en barometers (indicator 8). Bovendien passen de antwoordalternatieven in het algemeen goed bij de vraagtekst en zijn ze evenwichtig verdeeld over de *range* van mogelijke waarden van de variabele die met deze vraag gemeten wordt (indicatoren 6 en 7). En, tenslotte, bevatten de vragen duidelijke aanwijzingen omtrent beurtwisseling (indicator 12). De verschillen tussen de 37 vragen behorende tot het telefonische survey betreffen andere indicatoren dan de verschillen tussen de 9 vragen uit het 'face to face'-survey. In het telefonische survey betreffen de verschillen tussen de vragen onderling de indicatoren 1, 2, 3, 5, 10 en 11. In het 'face to face'-survey betreffen de verschillen tussen de 9 vragen onderling vooral de indicatoren 4, 6, 7, 8 en 10. Dat betekent dat ook in dit opzicht beide datasets elkaar goed aanvullen.

Geraadpleegde literatuur

Beatty, P. (1995) Understanding the standardized/non-standardized interviewing controversy. *Journal of Official Statistics* 11 (2) 147-160.

- Brenner, M. (1982) Response-effects of 'role-restricted' characteristics of the interviewer. In: W. Dijkstra en J. van der Zouwen (red.) *Response behaviour in the survey-interview*. London: Academic Press, p. 131-165.
- Dijkstra, W. (1994) SEQUENCE: A program for analysing sequential data. *Bulletin de Méthodologie Sociologique* 43: 134-142.
- Dijkstra, W. (te verschijnen) A new method for studying verbal interactions in survey-interviews. *Journal of Official Statistics*.
- Dijkstra, W., en J. van der Zouwen (1977) Testing auxiliary hypotheses behind the interview. *Annals of Systems Research* 6: 49-63.
- Dijkstra, W., en J. van der Zouwen (1988) Types of inadequate interviewer behaviour in survey-interviews. In: W.E. Saris en I.N. Gallhofer (red.) *Sociometric research. Volume 1: Data collection and scaling*. London: MacMillan, p. 24-35.
- Foddy, W. (1993) *Constructing questions for interviews and questionnaires*. Cambridge: Cambridge University Press
- Foddy, W. (1996) The in-depth testing of survey questions: A critical appraisal of methods. *Quality & Quantity* 30: 361-370.
- Houtkoop-Steenstra, H. (1995) Meeting both ends: Between standardization and recipient design in telephone survey interviews. In: P. ten Have en G. Psathas (red.) *Situated order: Studies in the social organization of talk and embodied activities*. Washington, D.C.: University Press of America, p. 91-106.
- Houtkoop-Steenstra, H. (1996) Probing behaviour of interviewers in the standardised semi-open research interview. *Quality & Quantity* 30: 205-230.
- Schaeffer, N.C., en D.W. Maynard (1996) From paradigm to prototype and back again: Interactive aspects of cognitive processing in standardized survey interviews. In: N. Schwarz en S. Sudman (red.) *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. San Francisco: Jossey-Bass, p. 65-88.
- Schwarz, N., en S. Sudman (red.) (1996) *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. San Francisco: Jossey-Bass.
- Sudman, S., en N.M. Bradburn (1982) *Asking questions: A practical guide to questionnaire design*. San Francisco: Jossey-Bass.
- Sykes, W., en M. Collins (1992) Anatomy of the survey interview. *Journal of Official Statistics* 8 (3) 277-291.
- Turner, C.F., en E. Martin (red.) (1984) *Surveying subjective phenomena*. New York: Russell Sage Foundation.
- Vaart, W. van der (1996) *Inquiring into the past: Data quality of responses to retrospective questions*. Proefschrift Vrije Universiteit, Amsterdam.
- Zouwen, J. van der (1974) A conceptual model for the auxiliary hypotheses behind the interview. *Annals of Systems Research* 4: 21-37.
- Zouwen, J. van der (1982) Hypotheses behind the sociological interview: Test and reformulation. In: R. Cavallo (red.) *Systems methodology in social science research*. Boston: Kluwer/Nijhoff, p. 142-157
- Zouwen, J. van der, en W. Dijkstra (1995) Trivial and non-trivial question-answer sequences: Types, determinants and effects on data quality. *Proceedings of the International Conference on Survey Measurement and Process Quality in Bristol (UK)*, p. 81 - 86.
- Zouwen, J. van der, en W. Dijkstra (1996) *The impact of the question on the interactions in survey-interviews*. Paper gepresenteerd tijdens de Fourth International ISA Conference on Social Science Methodology, July 1-5 University of Essex, Colchester UK.